

Network Working Group

M. Allman

Request for Comments: 2488

NASA Lewis/Sterling Software

BCP: 28

D. Glover

Category: Best Current Practice

NASA Lewis

L. Sanchez

BBN

January 1999

Повышение эффективности работы TCP через спутниковые каналы стандартными средствами

Enhancing TCP Over Satellite Channels using Standard Mechanisms

Статус документа

Этот документ относится к категории «Обмен опытом» (Best Current Practices) для сообщества Internet и служит приглашением к дискуссии в целях дальнейшего развития. Документ может распространяться свободно.

Авторские права

Copyright (C) The Internet Society (1999). All Rights Reserved.

Тезисы

Протокол TCP¹ обеспечивает гарантированную доставку данных по любому пути через сеть, включая пути со спутниковыми сегментами. Для работы TCP через спутниковые каналы имеется несколько стандартизованных IETF механизмов, позволяющих этому протоколу более эффективно использовать доступную полосу пропускания спутниковых каналов. В этом документе рассмотрены некоторые из этих «улучшений» TCP. В данный момент все рассматриваемые здесь механизмы имеют статус проектов стандартов IETF (или соответствуют стандартам IETF).

1. Введение

Спутниковые каналы могут оказывать влияние на работу транспортных протоколов типа TCP [Pos81]. При плохой работе протоколов (таких, как TCP) использование канала становится неэффективным. Хотя производительность транспортного протокола важна сама по себе, при работе по спутниковым каналам должны приниматься во внимание и другие аспекты. Например, протокол канального уровня, прикладные протоколы, размер буферов в маршрутизаторах, дисциплины очередей и расположение прокси могут оказывать значительное влияние на работу сети. Однако в этом документе рассматривается только повышение эффективности TCP в спутниковой среде, а остальные вопросы оставлены для других документов. Кроме того, предложены способы повышения эффективности передачи по спутниковым каналам на основе результатов исследований. Эти рекомендации могут быть полезны и безопасны для сетей совместного использования в будущем, однако данный документ рассматривает лишь механизмы TCP, которые уже хорошо изучены и включены в процессы стандартизации IETF (или совместимы со стандартами IETF).

Документ разделен на несколько частей: в разделе 2 приведены краткие характеристики спутниковых сетей, раздел 3 посвящен двум не относящимся к TCP механизмам, которые позволяют протоколу TCP более эффективно использовать доступную полосу каналов, в разделе 4 описаны определенные IETF механизмы TCP, которые могут быть полезны на спутниковых каналах, а в разделе 5 приведены рекомендации для современных реализаций TCP с точки зрения использования спутниковых каналов.

2. Характеристики спутниковых каналов

При доставке сообщений по спутниковым каналам неизбежно возникают задержки, обусловленные конечной величиной скорости света и большим удалением коммуникационных спутников.

Многие коммуникационные спутники размещаются на геостационарной орбите (GSO²), радиус которой составляет приблизительно 36 000 км [Sta94]. Нахождение спутника на геостационарной орбите обеспечивает его стабильное положение относительно земной поверхности. Т. е., каждая наземная станция всегда «видит» спутник в одной точке небосвода. Время распространения сигнала, определяемое двойным расстоянием до спутника составляет 239,6 мсек (для случая нахождения спутника строго вертикально над земной станцией) [Mag78]. Для наземных станций, находящихся на краях зоны обслуживания спутников расстояние составит 2 x 41 756 км, а задержка — 279,0 мсек [Mag78]. Приведенные значения задержек справедливы для одного «прыжка» (hop) земля-спутник-земля. Минимальное время между отправкой запроса и получением отклика на него (время кругового обхода - RTT³) будет составлять по крайней мере 558 мсек. Значение RTT обусловлено не только задержками на распространение радиосигнала между землей и спутником и находится под влиянием таких факторов, как задержки на передачу через наземные сети, ожидание в очередях шлюзов и т. п. Дополнительная задержка будет возникать в тех случаях, когда между конечными станциями будет более одного спутникового интервала. По мере реализации на спутниках операций по обработке сигналов это тоже может увеличивать задержки.

¹Transmission Control Protocol — протокол управления передачей.

²Geostationary Orbit.

³Round-trip time.

Спутники могут находиться и на других орбитах, включая LEO¹ [Stu95] [Mon98] и MEO² [Mar78]. Для связи через низкоорбитальные спутники последних требуется целая группировка. Иными словами, когда один спутник выходит из зоны видимости наземной станции, другой уже находится в этой зоне и «подхватывает» станцию. Задержка распространения сигналов для случая LEO составляет от нескольких миллисекунд для спутника, находящегося на одной вертикали с земной станцией, примерно до 80 мсек для спутника, приближающегося к горизонту. Такие системы зачастую используют межспутниковые соединения и задержка будет зависеть от маршрутизации в сети.

Два фундаментальных свойства спутниковых каналов рассмотрены ниже.

Шум — Уровень радиосигнала падает пропорционально квадрату пройденного расстояния. Для спутниковых каналов расстояние велико и сигнал ослабляется достаточно сильно. Это приводит к низкому отношению сигнал/шум. Некоторые частотные диапазоны весьма чувствительны к воздействиям атмосферы (например, затухание из-за дождей). Для мобильных приложений спутниковые каналы особенно чувствительны к возникающим из-за отражений множественным путям, а также к затенению (например, зданиями). Типичная частота битовых ошибок (BER³) для современного спутникового канала составляет не более 10^{-7} (1 ошибка на 10 миллионов битов). Эффективное кодирование с контролем ошибок (например, Reed Solomon) может быть реализовано на существующих спутниковых каналах и в настоящее время это уже используется во многих случаях. По мере внедрения новых механизмов контроля ошибок вероятность возникновения ошибок в спутниковых каналах становится сравнимой с вероятностью ошибок в оптических кабелях. Однако во многих старых системах значения BER во много раз превышают частоту ошибок в наземных каналах и новых спутниковых каналах.

Полоса пропускания — спектр радиочастот ограничен по своей природе, следовательно полоса спутниковых каналов тоже ограничена и обычно контролируется лицензиями. Дефицит частотного ресурса осложняет расширение полосы для решения других проблем. Для современных коммерческих спутниковых соединений «точка-точка» используется частота 6 ГГц в направлении спутника и 4 ГГц в направлении наземных станций (диапазон C) или 14/12 ГГц (диапазон Ku). Позднее был добавлен диапазон Ka с частотами 30/20 ГГц. Используемые на спутниках радио-повторители называют транспондерами. В диапазоне C полоса 36 МГц обычно обеспечивает передачу одного цветного телевизионного канала или 1200 телефонных каналов. Транспондеры диапазона Ku обычно работают с полосой канала 50 МГц. На одном спутнике может размещаться несколько десятков транспондеров.

Частотные диапазоны ограничены не только природой, но и международными соглашениями по распределению частот для коммерческой связи, поскольку частоты являются дефицитным ресурсом, требующимся для разных приложений.

Хотя спутниковые каналы по некоторым позициям проигрывают оптическим (сложность ремонта, атмосферные воздействия и т. п.), они обеспечивают и ряд преимуществ перед наземными каналами. Во-первых, спутниковые системы по своей природе являются широкоэмитерными. Это обеспечивает им преимущества в групповых приложениях. Кроме того, спутники позволяют обмениваться данными в труднодоступными территориями и странами, где нет развитой наземной инфраструктуры. Эффективно использование спутников и для мобильных приложений.

Спутниковые каналы имеют ряд характеристик, отличающих их от наземных каналов. К числу таких особенностей относится возможность снижения производительности TCP. Характеристики спутниковых каналов приведены ниже.

Long feedback loop — значительное время отклика

По причине продолжительной задержки в пути на некоторых спутниковых каналах (около 250 мсек для геостационарных спутников) отправителю TCP потребуются значительное время для того, чтобы убедиться в доставке пакета конечному получателю. Такие задержки осложняют работу интерактивных приложений типа telnet, а также работу некоторых механизмов контроля насыщения TCP (см. раздел 4).

Large delay*bandwidth product — большое значение задержка*полоса

Произведение задержки на пропускную способность канала (DBP⁴) определяет объем данных, которые протоколу следует рассматривать, как находящиеся «в пути» (переданы, но еще не подтверждены) для того, чтобы полностью использовать «емкость» канала. Используемое здесь значение задержки представляет собой время кругового обхода (RTT), а пропускная способность определяется самым узким местом на пути через сеть. Поскольку задержка на некоторых спутниковых каналах весьма велика, TCP требуется поддерживать большое число пакетов, находящихся «в пути» (переданы, но не подтверждены).

Transmission errors — ошибки при передаче

Число битовых ошибок (BER) на спутниковых каналах превышает число ошибок в типовых наземных каналах. TCP использует факты отбрасывания пакетов, как сигналы о перегрузке в сети и снижает размер окна в попытке преодолеть насыщение. При отсутствии данных о причине отбрасывания пакетов (перегрузка или повреждение) TCP должен предполагать, что причиной отбрасывания является перегрузка, чтобы избежать коллапса, описанного в [Jas88] [FF98]. Следовательно, отбрасывание поврежденных пакетов вынуждает TCP уменьшать размер окна даже в случае отсутствия реальных перегрузок в сети.

Asymmetric use - асимметрия

По причине высоких цен на передающее спутниковое оборудование в спутниковых сетях часто используются асимметричные решения. Например, подключенный к спутниковой сети хост передает весь исходящий трафик через медленный наземный канал (например, модем на коммутируемой линии), а входящий трафик принимает от спутника. В другом распространенном случае весь трафик передается через спутник, но для исходящего трафика используется меньшая полоса, нежели для входящего (по причине высокой цены широкополосного спутникового передатчика). Асимметрия может оказывать существенное влияние на работу TCP.

¹Low Earth Orbit — околоземная орбита.

²Medium Earth Orbit — средневисотная орбита.

³Bit error rate.

⁴Delay*bandwidth product.

Variable Round Trip Times — переменное значение RTT

В некоторых спутниковых системах (например, низкоорбитальных) задержка может существенно меняться с течением времени. Вопрос влияния таких изменений на работу TCP пока остается открытым.

Intermittent connectivity — прерывистая связность

В системах работающих не с геостационарными спутниками соединения TCP могут время от времени передаваться от одной группы спутников к другой или от одной наземной станции к другой. Такая передача может вызывать потерю пакетов.

Большинству спутниковых каналов присуща лишь часть перечисленных выше характеристик. Более того, перечисленные характеристики могут присутствовать не только в спутниковых средах. Тем не менее, для спутниковых сред перечисленные характеристики более характерны, а упомянутые проблемы проявляются сильнее. Отмеченные в этом документе механизмы должны обеспечить преимущества в большинстве случаев, особенно при наличии одной или нескольких упомянутых выше характеристик (например, в гигабитных сетях также наблюдаются большие значения $\text{delay} \cdot \text{bandwidth}$).

3. Улучшения на нижних уровнях

При использовании спутниковых каналов в сетях рекомендуется использовать два не связанных с TCP, которые могут повысить производительность TCP. Эти механизмы - Path MTU Discovery и FEC¹ - описаны в двух следующих параграфах.

Протокол канального уровня, используемый на спутниковом канале, может оказывать значительное влияние на работу протоколов вышележащих уровней. Хотя это и не входит в задачи данного документа, отметим, что при создании спутниковых сетей следует настраивать эти протоколы, чтобы они не ограничивали производительность работы TCP. В частности, протоколы канального уровня часто поддерживают механизмы управления окном потока данных и повтора передачи. Если размер окна на канальном уровне слишком мал, производительность тоже будет мала, как и при слишком малом размере окна TCP (подбор размера окна более подробно рассматривается в параграфе 4.3). Влияние повторов передачи на канальном уровне на производительность TCP в настоящее время исследовано не достаточно. Взаимодействие между механизмами повтора передачи на канальном уровне и в TCP требует дальнейшего изучения.

3.1 Определение Path MTU

Механизм Path MTU Discovery [MD90] служит для определения максимального размера пакетов, который соединение может использовать на данном пути через сеть без фрагментации IP. Отправитель передает пакет подходящего для локальной сети, к которой он подключен, размера (например, 1500 байтов для Ethernet) с установленным в заголовке флагом IP DF². Если пакет слишком велик для пересылки по сетевому пути без фрагментирования, шлюз который должен был фрагментировать пакет и переслать фрагменты дальше, вместо этого будет возвращать сообщение ICMP отправителю пакета. Это сообщение будет показывать невозможность передачи исходного пакета без фрагментирования и указывать максимальный размер пакета, который шлюз может переслать. Дополнительная информация IESG в части определения Path MTU приведена в [Kno93].

Path MTU Discovery позволяет TCP использовать максимально возможный размер пакетов без издержек на фрагментацию и сборку. Использование больших пакетов снижает издержки на заголовки. Как описано в разделе 4, увеличение размера окна TCP основано на размере сегментов, поэтому большие сегменты TCP позволяют отправителям быстрее увеличивать окно насыщения.

Однако результатом применения Path MTU Discovery может быть задержка перед тем, как TCP сможет начать передачу данных. Предположим, например, что передается пакет с установленным флагом DF и один из промежуточных шлюзов (G1) возвращает сообщение ICMP указывающее невозможность передачи без фрагментирования. После этого хост снижает размер пакетов до значения, указанного в сообщении ICMP от шлюза G1 и передает новый пакет с флагом DF. Этот пакет будет переслан шлюзом G1, однако это не дает гарантий того, что все последующие шлюзы смогут передать этот сегмент без фрагментирования. Если другой шлюз (G2) не может переслать сегмент, он будет возвращать отправителю сообщение ICMP и процесс повторится. Следовательно определение MTU для пути может потребовать достаточно большого времени. Задержки на спутниковых каналах будут усугублять проблему (например, спутниковый канал между G1 и G2). Однако на практике Path MTU Discovery не занимает чрезмерного времени, поскольку значения MTU достаточно «стандартны». Кроме того, кэширование значений MTU во многих случаях позволяет смягчить проблему, хотя реализация такого кэширования и время хранения записей в кэше требуют дополнительного изучения.

Связь между значением BER и размером сегментов существенно зависит от характеристик конкретного канала. Эти связи требуют дальнейших исследований, однако при использовании упреждающей коррекции ошибок (параграф 3.2) увеличение размера сегментов будет повышать производительность как и в других сетях [MSMO97]. Хотя точный метод выбора лучшего значения MTU для спутниковых каналов выходит за рамки этого документа, рекомендуется применять Path MTU Discovery, чтобы протокол TCP мог использовать максимально возможное значение MTU на спутниковом канале.

3.2 Упреждающая коррекция ошибок

Потери в TCP всегда трактуются, как индикация перегрузки, и вынуждают TCP уменьшить размер окна насыщения. Поскольку рост этого окна основан на возврате подтверждений (см. раздел 4), восстановление TCP после потери занимает достаточно много времени в случае работы по спутниковым каналам. Если потеря пакета связана с повреждением, а не отбрасыванием в результате перегрузки, TCP не требуется снижать размер окна насыщения. Однако к настоящему времени детектирование потерь, связанных с повреждением исследовано не достаточно.

Следовательно, для эффективной работы TCP требуется обеспечить такие характеристики канала, чтобы не возникало потерь пакетов, связанных с их повреждением. Следует применять упреждающую коррекцию ошибок (FEC) для улучшения BER на спутниковых каналах. Однако в спутниковой среде снижение BER возможно не всегда. Поскольку восстановление TCP после потери пакетов занимает продолжительное время по причине большой задержки

¹Forward error correction — упреждающая корректировка ошибок.

²Don't fragment — не фрагментировать.

распространения сигналов в спутниковом канале [PS97], следует обеспечивать максимальную «чистоту» канала для предотвращения ложных сигналов о перегрузке соединений TCP. В этом документе рекомендуется лишь максимально снижать значение BER для повышения эффективности работы TCP.

Не следует ожидать от FEC решения всех проблем, связанных со спутниковым каналом. Есть несколько ситуаций, где FEC не может решить проблему шума (например, военные помехи, дальняя космическая связь, тропический ливень). Кроме того, перебои связи могут быть связаны с самими спутниковыми системами, хотя это случается реже, чем в наземных устройствах. И, наконец, FEC не дается даром. Для поддержки FEC требуется дополнительное оборудование и расходуется часть полосы канала. Коррекция ошибок может также увеличивать задержку и ее вариации за счет затраты времени на кодирование и декодирование.

Требуются дальнейшие исследования в части механизмов, позволяющих протоколу TCP различать потери в результате перегрузок и повреждения пакетов. Такие механизмы позволят TCP отвечать подобающим способом на перегрузку и пытаться исправить поврежденные пакеты без снижения скорости передачи. Однако в отсутствие таких механизмов потеря пакета должна для предотвращения нестабильности в сети восприниматься, как индикация перегрузки. Некорректная интерпретация потери, как результата повреждения пакета, без снижения скорости передачи может привести к коллаптической перегрузке [Jac88] [FF98].

4. Стандартные механизмы TCP

В этом разделе рассмотрены механизмы TCP, которые могут потребоваться при работе по спутниковым или гибридным (земля-спутник) каналам, для повышения эффективности использования доступной полосы. Эти же механизмы могут помочь повысить эффективность пропускную способность наземных каналов. Кроме того, эти механизмы не мешают работе наземных сетей общего пользования. В каждом из последующих параграфов кратко рассмотрен один из механизмов и описаны причины его использования.

4.1 Контроль насыщения

Для предотвращения генерации неприемлемого в текущих условиях сетевого трафика на соединениях TCP реализуется четыре механизма контроля перегрузок [Jac88] [Jac90] [Ste97] — замедленный старт (slow start), предотвращение насыщения (congestion avoidance), ускоренный повтор передачи (fast retransmit) и быстрое восстановление (fast recovery). Эти алгоритмы контролируют объем неподтвержденных данных, которые могут быть переданы, и повтор передачи отброшенных в сети пакетов.

Отправители TCP используют две переменных состояния в целях контроля перегрузок. Первой переменной является окно насыщения $swnd$ ¹. Это значение определяет верхнюю границу объема данных, которые отправитель может передать в сеть до того, как получит подтверждение (ACK). Значение $swnd$ ограничено размером анонсируемого получателем окна. Размер окна насыщения может меняться в процессе передачи на основе заключений об уровне перегрузки в сети. Второй переменной является порог замедленного старта $ssthresh$ ². Эта переменная определяет алгоритм, используемый для управления значением $swnd$. Если $swnd < ssthresh$, используется алгоритм замедленного старта (slow start) для увеличения $swnd$. Однако при $swnd \geq ssthresh$ (или просто $>$ для некоторых реализаций TCP), применяется алгоритм предотвращения перегрузки. Начальное значение $ssthresh$ совпадает с анонсируемым получателем размером окна. Кроме того, значение $ssthresh$ устанавливается при обнаружении перегрузки.

Четыре алгоритма контроля перегрузок рассматриваются ниже, а после этого кратко обсуждается влияние спутниковой среды на работу этих алгоритмов.

4.1.1 Замедленный старт и предотвращение перегрузок

Начиная передачу через соединение TCP, хост не знает состояния сети между собой и получателем данных. Для предотвращения отправки неприемлемо большого объема данных отправителю нужно в начале передачи использовать алгоритм замедленного старта [Jac88] [Bra89] [Ste97]. Процедура начинается с инициализации $swnd = 1$ (хотя экспериментальный механизм IETF будет увеличивать размер начального окна приблизительно до 4 Кбайт [AFP98]) и установки для $ssthresh$ размера анонсируемого получателем окна. Это заставляет TCP передать один сегмент и ждать соответствующего подтверждения ACK. Для каждого ACK, полученного во время процедуры замедленного старта, значение $swnd$ увеличивается на 1 сегмент. Например, после первого принятого ACK будет установлено $swnd = 2$, что позволяет отправителю передать 2 пакета данных. Процедура повторяется до тех пор, пока $swnd$ достигнет или превзойдет $ssthresh$ (или, в некоторых реализациях, $swnd = ssthresh$) или не возникнет потеря.

Когда значение $swnd$ достигает или превышает (или равно для некоторых реализаций) $ssthresh$, используется алгоритм предотвращения перегрузок для увеличения [Jac88] [Bra89] [Ste97]. Это алгоритм обеспечивает более медленное увеличение размера окна по сравнению с процедурой замедленного старта. Предотвращение перегрузок используется для медленного зондирования пропускной способности сети. В процессе congestion avoidance размер $swnd$ увеличивается на $1/swnd$ по каждому принятому подтверждению ACK. Следовательно, если приходит одно подтверждение для каждого сегмента данных, $swnd$ будет увеличиваться примерно на 1 сегмент за период кругового обхода (RTT).

Алгоритмы замедленного старта и предотвращения перегрузок могут снижать эффективную пропускную способность канала при работы в спутниковых сетях с большими задержками [All97]. Например, передача начинается с окна размером 1 сегмент. После его передачи отправитель должен ждать приема соответствующего пакета ACK. На спутниковых каналах GSO время ожидания составит приблизительно 500 мсек и в течение этого времени соединение будет простаивать. Процедура замедленного старта на спутниковых каналах GSO будет длиться значительно дольше, чем на типичных наземных каналах. Это задержит начало использования алгоритма предотвращения перегрузок [All97]. По этой причине так важно использование механизмов Path MTU Discovery. Хотя число сегментов все равно будет определяться алгоритмами контроля насыщения, размеры сегмента зависят от MTU.

Следовательно, использование более крупных пакетов позволит TCP передавать в сегменте больше данных и повысит эффективность использования канала.

¹Congestion window.

²Slow start threshold.

4.1.2 Ускоренный повтор и быстрое восстановление

По умолчанию TCP использует для детектирования отброшенных сегментов тайм-аут [Pos81]. Иными словами, если отправитель не получит ACK для данного пакета за ожидаемый интервал времени, передача сегмента повторяется. Тайм-аут повторной передачи (RTO¹) определяется на основе наблюдений за RTT. В дополнение к повтору передачи по истечении RTO протокол TCP использует факт потери сегмента в качестве индикации насыщения в сети. В ответ на перегрузку значение ssthresh устанавливается равным половине cwnd, а затем значение cwnd уменьшается до 1 сегмента. Это инициирует подключение алгоритма замедленного старта для увеличения cwnd, пока оно не достигнет половины значения cwnd на момент обнаружения перегрузки. После фазы замедленного старта используется алгоритм предотвращения перегрузок для проверки наличия дополнительной пропускной способности в сети.

Пакеты TCP ACK всегда повторяют пакет со старшим номером из числа доставленных с сохранением порядка. Следовательно, ACK для сегмента X также является подтверждением для всех сегментов с номерами меньше X. Более того, при доставке сегмента с нарушением порядка пакет ACK все равно будет подтверждать старший номер доставленных упорядоченно сегментов, а не прибывший сегмент. Например, предположим, что сегмент 11 был отброшен в сети, а сегмент 12 принят получателем. Получатель в этом случае передаст дубликат ACK, подтверждающий доставку сегмента 10 (и всех предшествующих).

Алгоритм ускоренного повтора использует такие дубликаты ACK для детектирования потери сегментов. Если отправитель данных получил 3 дубликата ACK, TCP предполагает потерю сегмента и повторяет передачу отсутствующего сегмента, не ожидая тайм-аута RTO. После отправки сегмента с использованием ускоренного повтора используется алгоритм быстрого восстановления для настройки окна насыщения. Сначала устанавливается значение ssthresh, равное половине cwnd. Затем значение cwnd уменьшается наполовину. В заключение cwnd дополнительно увеличивается на 1 для каждого полученного дубликата ACK. Такое увеличение обусловлено тем, что каждый дубликат ACK представляет уже покинувший сеть сегмент. Если cwnd позволяет, TCP может передать новые данные. Такой подход позволяет TCP сохранять поток данных на уровне половины скорости, при которой была обнаружена потеря пакетов. При получении ACK для переданных повторно пакетов значение cwnd снижается обратно до ssthresh (половина значения cwnd в момент обнаружения перегрузки).

В общем случае при ускоренном повторе может быть повторен только один сегмент на окно переданных данных. При потере множества сегментов из одного окна один из них будет передан заново с использованием ускоренного повтора, а для остальных потерянных сегментов придется, как обычно, ждать тайм-аута RTO, по которому TCP инициирует процедуру slow start.

Отклик TCP на перегрузку зависит от способа ее обнаружения. Если повторная передача пакета была вызвана таймером повтора, TCP сбрасывает значение ssthresh до половины текущего значения cwnd и снижает значение cwnd до 1 сегмента (иницируя, таким образом, замедленный старт). Однако, если повтор передачи был вызван механизмом fast retransmit для обоих параметров ssthresh и cwnd значения уменьшаются до половины текущего значения cwnd и для отправки новых данных используется механизм предотвращения перегрузок. Различие заключается в том, что при повторе, вызванном дубликатами ACK, TCP знает, что пакеты продолжают находиться в сети и можно считать, что перегрузка не велика. Однако при повторе передачи по таймеру повтора TCP ничего не знает о состоянии сети и, следовательно, должен быть консервативным при передаче новых данных, используя механизм замедленного старта.

Отметим, что описанное выше применение алгоритмов повтора и быстрого восстановления может приводить к возникновению множества быстрых повторов передачи в одном окне данных [Flo94]. Это может привести к многократному снижению размера окна насыщения в ответ на один факт потери пакета. Проблема особенно заметна для соединений с большим окном насыщения, поскольку такие соединения могут отправлять в сеть достаточно много новых сегментов в процессе восстановления для запуска множества ускоренных повторов. Многократное снижение значения cwnd в ответ на потерю одного пакета может существенно снизить производительность [GJKFV98].

Лучшим способом повысить эффективность механизмов ускоренного повтора и восстановления является детектирование потерь на основе механизма селективных подтверждений. Как описано ниже, такие механизмы обычно способны обеспечить быстрое восстановление при потере множества сегментов без необходимости снижения cwnd. В отсутствие SACK следует применять алгоритмы ускоренного повтора и восстановления. Корректировка этих алгоритмов для достижения высокой производительности в случае множественных ускоренных повторов выходит за рамки данного документа. По этой причине разработчикам TCP следует реализовать текущие варианты алгоритмов ускоренного повтора и быстрого восстановления, описанных в RFC 2001 [Ste97] или последующих вариантах RFC 2001.

4.1.3 Контроль насыщения в спутниковой среде

Упомянутые выше алгоритмы оказывают негативное влияние на производительность отдельных соединений TCP, поскольку алгоритмы достаточно зондируют сеть на предмет дополнительных «мощностей», потребляя для этого некоторую часть доступной пропускной способности. Это, в частности, правильно для спутниковых каналов со значительными задержками, поскольку на получение отправителем отклика от получателя уходит много времени [All97] [АНКО97]. Однако эти алгоритмы необходимы для предотвращения коллапса насыщения в сети с совместным доступом [Jas88]. Следовательно, негативное влияние на отдельные соединения приходится принимать ради преимуществ сети в целом.

4.2 Большое окно TCP

Стандартное значение максимального размера окна TCP (65 535 байтов) не позволяет одному соединению TCP полностью воспользоваться доступной полосой на некоторых спутниковых каналах. Пропускная способность соединения TCP определяется формулой [Pos81]:

$$\text{throughput} = \text{window size} / \text{RTT}$$

Следовательно, при максимальном размере окна 65 535 байтов и работе через геостационарный спутник с RTT = 560 мсек [Kru95] максимальная пропускная способность составит

$$\text{throughput} = 65535 / 560 = 117027 \text{ байт/сек.}$$

Следовательно, одно стандартное соединение TCP не сможет полностью воспользоваться, например, каналом T1 (скорость около 192000 байт/сек.) при работе через спутник GSO. Однако протокол TCP был расширен с целью

¹Retransmission timeout.

поддержки больших размеров окон [JBB92]. Описанное в [JBB92] масштабирование окна следует применять в спутниковых средах вместе с алгоритмами PAWS¹ и RTTM².

Следует отметить, что при использовании одного спутникового канала для множества соединений увеличение размера окна не является обязательным. Например, два долгосрочных соединения TCP с окнами размером 65535 байтов могут полностью использовать канал T1 через спутник GSO.

Применение больших окон зачастую требует ручной настройки клиентских и серверных приложений или стеков TCP (обычно требуется опытный специалист). Ведутся исследования механизмов операционных систем для настройки буферной емкости в соответствии с текущими условиями в сети [SMM98]. Результаты таких исследований обеспечат реализациям TCP и приложениям возможность более эффективного использования обеспечиваемой нижележащими уровнями производительности.

4.3 Стратегия подтверждений

Есть два стандартных метода генерации подтверждений получателями TCP. В методе [Pos81] генерируется пакет ACK для каждого входящего сегмента. В работе [Bra89] утверждается, что хостам **следует** использовать отложенные подтверждения. При использовании этого алгоритма пакеты ACK генерируются для каждого второго полноразмерного сегмента или, если второй полноразмерный сегмент не приходит в течение заданного времени (обычно не более 500 мсек.). Окно насыщения увеличивается в соответствии с номерами входящих пакетов ACK и отложенная отправка подтверждений снижает число пакетов ACK, отправленных получателем. Следовательно, рост параметра cwnd замедляется при использовании отложенных подтверждений по сравнению со случаем генерации ACK на каждый принятый сегмент [All98].

Заманчивым «решением» проблемы, вызываемой отложенными подтверждениями, кажется отключение данного механизма и передача ACK для каждого входящего сегмента. Однако делать это не рекомендуется. Во-первых, в документе [Bra89] указано, что получателям TCP **следует** использовать отложенные подтверждения. Во-вторых, двойное увеличение числа пакетов ACK в разделяемой сети может привести к неочевидным последствиям. Следовательно, запрет использования отложенных подтверждений требует дополнительного изучения и в настоящее время получателям TCP следует продолжать генерацию пакетов ACK в соответствии с [Bra89].

4.4 Селективные подтверждения

Селективные подтверждения (SACK³) [MMFR96] позволяют получателям TCP сообщать отправителям точную информацию о доставленных пакетах. SACK позволяют TCP быстрее восстанавливаться после потери сегментов, а также предотвращают ненужные повторы передачи.

Алгоритм ускоренного повтора обычно позволяет исправить лишь одну потерю на окно данных. При возникновении множества потерь отправитель обычно должен использовать тайм-аут для определения сегмента, который нужно повторно передать следующим. В процессе ожидания тайм-аута сегменты данных и подтверждения для них уходят из сети. При отсутствии входящих пакетов ACK для инициирования отправки в сеть новых сегментов отправитель должен использовать для повтора передачи алгоритм замедленного старта. Как отмечено выше, применение этого алгоритма на спутниковых каналах может быть связано со значительными временными издержками. При использовании SACK отправитель обычно может определить какие сегменты следует передать повторно за первый период RTT после обнаружения потери. Это позволяет отправителю продолжать передачу сегментов (отправлять новые сегменты, если это приемлемо) с подходящей скоростью и, следовательно, поддерживать синхронизацию ACK. Это позволяет избавиться от длительной процедуры замедленного старта при потере множества сегментов. Обычно применение SACK позволяет повторить передачу всех потерянных сегментов в течение первого периода RTT после обнаружения потери. В работах [MM96] и [FF96] рассмотрены специфические механизмы контроля насыщения, использующие информацию SACK для определения необходимости повторной передачи и подходящего для повтора времени. В обоих этих алгоритмах используются базовые принципы контроля насыщения, описанные в [Jas88] и при обнаружении перегрузки окно уменьшается наполовину.

5. Обзор улучшений

В таблице 1 перечислены рассмотренные в этом документе механизмы. Некоторые механизмы имеют статус Recommended в процессе стандартизации IETF и рекомендуются авторами для использования в сетях со спутниковыми каналами. Часть механизмов обязательна для применения (статус Required в процессе стандартизации IETF) и они должны использоваться на хостах с доступом в Internet [Bra89]. В таблице указаны разделы документа, посвященные каждому механизму и места использования — символ S указывает механизмы, используемые на стороне отправителя, R — на стороне получателя и L — на спутниковых каналах.

Таблица 1

Механизм	Применение	Описание	Место
Path-MTU Discovery	Рекомендуется	3.1	S
FEC	Рекомендуется	3.2	L
TCP Congestion Control			
Slow Start	Обязательно	4.1.1	S
Congestion Avoidance	Обязательно	4.1.1	S
Fast Retransmit	Рекомендуется	4.1.2	S
Fast Recovery	Рекомендуется	4.1.2	S

¹Protection Against Wrapped Sequence space — защита от перехода через границу отсчета порядковых номеров.

²Round-Trip Time Measurements — измерение времени кругового обхода.

³Selective acknowledgment.

TCP Large Windows			
Window Scaling	Рекомендуется	4.2	S, R
PAWS	Рекомендуется	4.2	S, R
RTTM	Рекомендуется	4.2	S, R
TCP SACK	Рекомендуется	4.4	S, R

Пользователям спутниковых каналов следует выяснить у разработчиков (производителей), поддерживают ли их реализации TCP рекомендованные здесь механизмы. Pittsburgh Supercomputer Center отслеживает функциональность реализаций TCP и поддерживаемых ими расширений, а также публикует рекомендации по настройке различных реализаций TCP [PSC].

Исследования по оптимизации эффективности TCP на спутниковых каналах продолжаются и результаты будут опубликованы вместе с обсуждением других вопросов типа архитектуры спутниковых сетей..

6. Вопросы безопасности

Авторы надеются, что приведенные здесь рекомендации не окажут влияния на безопасность TCP. Однако при использовании широкоэмитательных сред типа спутниковых каналов для передачи пользовательских данных и/или трафика сетевого управления следует разобраться с вопросами внутренней безопасности применяемых технологий.

Перехват в сетевых каналах является одной из форм пассивных атак и при его успешной реализации может привести к утечке важных данных управления сетью, что может послужить причиной последующего нарушения работы сети. Такие атаки могут снижать эффективность передачи данных через сети. Перехват данных может также приводить к утечке конфиденциальной информации пользователей особенно в тех случаях, когда не применяется механизмов сквозной защиты. Пассивный мониторинг возможен в любой сети, однако широкоэмитательная природа спутниковых каналов позволяет организовать перехват без физического подключения к сети и обнаружение такого перехвата весьма затруднительно. Следует однако отметить, что требуемые для такого перехвата ресурсы делают его нетривиальной задачей.

Шифрование данных на физическом и/или канальном уровне может обеспечить защиту данных, передаваемых по спутниковым каналам. Однако это не избавляет от возможности перехвата данных до или после спутникового канала. В этом документе не дается каких-либо рекомендаций по выбору механизмов защиты. Однако пользователям спутниковых каналов следует рассмотреть доступные механизмы защиты сетей и выбрать среди них подходящие.

Благодарности

При подготовке этого документа важную роль сыграли комментарии членов рабочей группы TCP Over Satellite. В частности, авторы благодарят Aaron Falk, Matthew Halsey, Hans Kruse, Matt Mathis, Greg Nakanishi, Vern Paxson, Jeff Semke, Bill Serpeier и Eric Travis за их полезные замечания к данному документу.

Литература

- [AFP98] Allman, M., Floyd, S. and C. Partridge, "Increasing TCP's Initial Window", RFC 2414¹, September 1998.
- [АНКО97] Mark Allman, Chris Hayes, Hans Kruse, and Shawn Ostermann. TCP Performance Over Satellite Links. In Proceedings of the 5th International Conference on Telecommunication Systems, March 1997.
- [All97] Mark Allman. Improving TCP Performance Over Satellite Channels. Master's thesis, Ohio University, June 1997.
- [All98] Mark Allman. On the Generation and Use of TCP Acknowledgments. ACM Computer Communication Review, 28(5), October 1998.
- [Bra89] Braden, R., "Requirements for Internet Hosts – Communication Layers", STD 3, [RFC 1122](#), October 1989.
- [FF96] Kevin Fall and Sally Floyd. Simulation-based Comparisons of Tahoe, Reno and SACK TCP. Computer Communication Review, July 1996.
- [FF98] Floyd, Kevin Fall. Promoting the Use of End-to-End Congestion Control in the Internet. Submitted to IEEE Transactions on Networking.
- [Flo94] S. Floyd, TCP and Successive Fast Retransmits. Technical report, October 1994. <ftp://ftp.ee.lbl.gov/papers/fastretrans.ps>.
- [GJKFV98] Rohit Goyal, Raj Jain, Shiv Kalyanaraman, Sonia Fahmy, Bobby Vandalore, Improving the Performance of TCP over the ATM-UBR service, 1998. Submitted to Computer Communications.
- [Jac90] Van Jacobson. Modified TCP Congestion Avoidance Algorithm. Technical Report, LBL, April 1990.
- [JBB92] Jacobson, V., Braden, R. and D. Borman, "TCP Extensions for High Performance", [RFC 1323](#), May 1992.
- [Jac88] Van Jacobson. Congestion Avoidance and Control. In ACM SIGCOMM, 1988.
- [Kno93] Knowles, S., "IESG Advice from Experience with Path MTU Discovery", RFC 1435, March 1993.
- [Mar78] James Martin. Communications Satellite Systems. Prentice Hall, 1978.
- [MD90] Mogul, J. and S. Deering, "Path MTU Discovery", [RFC 1191](#), November 1990.
- [MM96] Matt Mathis and Jamshid Mahdavi. Forward Acknowledgment: Refining TCP Congestion Control. In ACM SIGCOMM, 1996.
- [MMFR96] Mathis, M., Mahdavi, J., Floyd, S. and A. Romanow, "TCP Selective Acknowledgment Options", [RFC 2018](#), October 1996.

¹Этот документ признан устаревшим и заменен [RFC 3390](#).

- [Mon98] M. J. Montpetit. TELEDESIC: Enabling The Global Community Interaccess. In Proc. of the International Wireless Symposium, May 1998.
- [MSMO97] M. Mathis, J. Semke, J. Mahdavi, T. Ott, "The Macroscopic Behavior of the TCP Congestion Avoidance Algorithm", Computer Communication Review, volume 27, number3, July 1997. available from <http://www.psc.edu/networking/papers/papers.html>.
- [Pos81] Postel, J., "Transmission Control Protocol", STD 7, [RFC 793](#), September 1981.
- [PS97] Craig Partridge and Tim Shepard. TCP Performance Over Satellite Links. IEEE Network, 11(5), September/October 1997.
- [PSC] Jamshid Mahdavi. Enabling High Performance Data Transfers on Hosts. http://www.psc.edu/networking/perf_tune.html.
- [SMM98] Jeff Semke, Jamshid Mahdavi and Matt Mathis. Automatic TCP Buffer Tuning. In ACM SIGCOMM, August 1998. To appear.
- [Sta94] William Stallings. Data and Computer Communications. MacMillan, 4th edition, 1994.
- [Ste97] Stevens, W., "TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms", [RFC 2001](#), January 1997.
- [Stu95] M. A. Sturza. Architecture of the TELEDESIC Satellite System. In Proceedings of the International Mobile Satellite Conference, 1995.

Адреса авторов

Mark Allman

NASA Lewis Research Center/Sterling Software
21000 Brookpark Rd. MS 54-2
Cleveland, OH 44135
Phone: +1 216 433 6586
E-Mail: mallman@lerc.nasa.gov
<http://roland.lerc.nasa.gov/~mallman>

Daniel R. Glover

NASA Lewis Research Center
21000 Brookpark Rd.
Cleveland, OH 44135
Phone: +1 216 433 2847
E-Mail: Daniel.R.Glover@lerc.nasa.gov

Luis A. Sanchez

BBN Technologies
GTE Internetworking
10 Moulton Street
Cambridge, MA 02140
USA
Phone: +1 617 873 3351
E-Mail: lsanchez@ir.bbn.com

Перевод на русский язык

Николай Малых
nmalykh@gmail.com

Полное заявление авторских прав

Copyright (C) The Internet Society (1999). All Rights Reserved.

This document and translations of it may be copied and furnished to others, and derivative works that comment on or otherwise explain it or assist in its implementation may be prepared, copied, published and distributed, in whole or in part, without restriction of any kind, provided that the above copyright notice and this paragraph are included on all such copies and derivative works. However, this document itself may not be modified in any way, such as by removing the copyright notice or

references to the Internet Society or other Internet organizations, except as needed for the purpose of developing Internet standards in which case the procedures for copyrights defined in the Internet Standards process must be followed, or as required to translate it into languages other than English.

The limited permissions granted above are perpetual and will not be revoked by the Internet Society or its successors or assigns.

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.